

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 2 年 1 2 月 1 2 日
Date of Application:

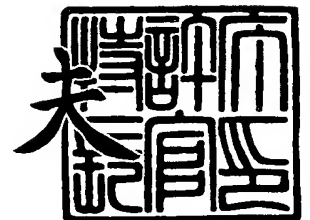
出 願 番 号 特 願 2 0 0 2 - 3 6 1 3 0 2
Application Number:
[ST. 10/C]: [J P 2 0 0 2 - 3 6 1 3 0 2]

出 願 人 ソニー株式会社
Applicant(s):

2 0 0 3 年 9 月 8 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康



【書類名】 特許願

【整理番号】 0290792706

【提出日】 平成14年12月12日

【あて先】 特許庁長官 太田 信一郎 殿

【国際特許分類】 G11B 27/00
G06F 17/30
G10L 15/02

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 安部 素嗣

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 向井 昭広

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 西口 正之

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代理人】

【識別番号】 100067736

【弁理士】

【氏名又は名称】 小池 晃

【選任した代理人】

【識別番号】 100086335

【弁理士】

【氏名又は名称】 田村 榮一

【選任した代理人】

【識別番号】 100096677

【弁理士】

【氏名又は名称】 伊賀 誠司

【手数料の表示】

【予納台帳番号】 019530

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9707387

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音響信号処理装置及び方法、信号記録装置及び方法、並びにプログラム

【特許請求の範囲】

【請求項 1】 音響信号における特徴部分を検出する音響信号処理装置において、

上記音響信号の所定時間区間毎に短時間振幅を計算する振幅計算手段と、

上記短時間振幅に基づいて上記特徴部分の候補区間を抽出する候補区間抽出手段と、

上記音響信号の所定時間区間毎に音質を定量化する音質特徴量を抽出する特徴抽出手段と、

上記短時間振幅及び上記音質特徴量に基づいて、上記候補区間が上記特徴部分か否かを評価する候補区間評価手段と

を備えることを特徴とする音響信号処理装置。

【請求項 2】 上記音響信号は、イベント中に発生する音響信号であることを特徴とする請求項 1 記載の音響信号処理装置。

【請求項 3】 上記特徴部分の少なくとも開始位置及び終了位置を含む索引情報を生成する索引生成手段をさらに備えることを特徴とする請求項 1 記載の音響信号処理装置。

【請求項 4】 上記音質特徴量は、上記音響信号の短時間パワースペクトラム係数、上記音響信号の短時間ケプストラム係数、又は上記短時間パワースペクトラム係数にカルーネンレーベ変換を施した係数であることを特徴とする請求項 1 記載の音響信号処理装置。

【請求項 5】 上記候補区間抽出手段は、上記短時間振幅が第 1 の閾値を超える区間が第 2 の閾値を超えて継続する場合に、該区間を候補区間として抽出することを特徴とする請求項 1 記載の音響信号処理装置。

【請求項 6】 上記候補区間評価手段は、上記候補区間の区間長、上記候補区間における上記短時間振幅の最大値若しくは平均値、上記候補区間における上記音質特徴量が所定の閾値を超える長さ若しくはその候補区間長に対する割合、又

は上記候補区間における上記音質特徴量の最大値若しくは平均値の少なくとも 1 つを用いて、上記候補区間が上記特徴部分か否かを評価することを特徴とする請求項 1 記載の音響信号処理装置。

【請求項 7】 上記音響信号における挿入信号を検出する挿入検出手段をさらに備え、

上記候補区間評価手段は、上記挿入検出手段によって検出された挿入区間に基づいて、上記候補区間が上記特徴部分か否かを評価する

ことを特徴とする請求項 1 記載の音響信号処理装置。

【請求項 8】 上記挿入検出手段は、上記短時間振幅が所定の閾値未満となる小音量区間が所定時間内に複数回発生した場合に、この小音量区間によって挟まれた区間を上記挿入区間として検出することを特徴とする請求項 7 記載の音響信号処理装置。

【請求項 9】 上記音響信号は、コマーシャルメッセージを含むものであり、上記挿入検出手段は、上記コマーシャルメッセージを上記挿入信号として検出することを特徴とする請求項 7 記載の音響信号処理装置。

【請求項 10】 音響信号における特徴部分を検出する音響信号処理方法において、

上記音響信号の所定時間区間毎に短時間振幅を計算する振幅計算工程と、

上記短時間振幅に基づいて上記特徴部分の候補区間を抽出する候補区間抽出工程と、

上記音響信号の所定時間区間毎に音質を定量化する音質特徴量を抽出する特徴抽出工程と、

上記短時間振幅及び上記音質特徴量に基づいて、上記候補区間が上記特徴部分か否かを評価する候補区間評価工程と

を有することを特徴とする音響信号処理方法。

【請求項 11】 上記音響信号は、イベント中に発生する音響信号であることを特徴とする請求項 10 記載の音響信号処理方法。

【請求項 12】 上記特徴部分の少なくとも開始位置及び終了位置を含む索引情報を生成する索引生成工程をさらに有することを特徴とする請求項 10 記載の

音響信号処理方法。

【請求項 1 3】 音響信号における特徴部分を検出する処理をコンピュータに実行させるプログラムにおいて、

上記音響信号の所定時間区間毎に短時間振幅を計算する振幅計算工程と、

上記短時間振幅に基づいて上記特徴部分の候補区間を抽出する候補区間抽出工程と、

上記音響信号の所定時間区間毎に音質を定量化する音質特徴量を抽出する特徴抽出工程と、

上記短時間振幅及び上記音質特徴量に基づいて、上記候補区間が上記特徴部分か否かを評価する候補区間評価工程と

を有することを特徴とするプログラム。

【請求項 1 4】 上記音響信号は、イベント中に発生する音響信号であることを特徴とする請求項 1 3 記載のプログラム。

【請求項 1 5】 上記特徴部分の少なくとも開始位置及び終了位置を含む索引情報を生成する索引生成工程をさらに有することを特徴とする請求項 1 3 記載のプログラム。

【請求項 1 6】 音響信号の所定時間区間毎に短時間振幅を計算する振幅計算手段と、

上記短時間振幅に基づいて上記音響信号の重要部分の候補区間を抽出する候補区間抽出手段と、

上記音響信号の所定時間区間毎に音質を定量化する音質特徴量を抽出する特徴抽出手段と、

上記短時間振幅及び上記音質特徴量に基づいて上記候補区間の重要度を計算し、該重要度に基づいて上記候補区間が上記重要部分か否かを評価する候補区間評価手段と、

上記候補区間評価手段によって上記重要部分であると評価された候補区間の少なくとも開始位置、終了位置及び重要度を含む索引情報を生成する索引生成手段と、

上記音響信号と共に上記索引情報が記録される記録手段と

を備えることを特徴とする信号記録装置。

【請求項 1 7】 上記音響信号は、イベント中に発生する音響信号であることを特徴とする請求項 1 6 記載の信号記録装置。

【請求項 1 8】 上記音質特徴量は、上記音響信号の短時間パワースペクトラム係数、上記音響信号の短時間ケプストラム係数、又は上記短時間パワースペクトラム係数にカルーネンレーベ変換を施した係数であることを特徴とする請求項 1 6 記載の信号記録装置。

【請求項 1 9】 上記候補区間抽出手段は、上記短時間振幅が第 1 の閾値を超える区間が第 2 の閾値を超えて継続する場合に、該区間を候補区間として抽出することを特徴とする請求項 1 6 記載の信号記録装置。

【請求項 2 0】 上記候補区間評価手段は、上記候補区間の区間長、上記候補区間における上記短時間振幅の最大値若しくは平均値、上記候補区間における上記音質特徴量が所定の閾値を超える長さ若しくはその候補区間長に対する割合、又は上記候補区間における上記音質特徴量の最大値若しくは平均値の少なくとも 1 つを用いて、上記候補区間の重要度を計算することを特徴とする請求項 1 6 記載の信号記録装置。

【請求項 2 1】 上記音響信号が適用可能なジャンルであるか否かを識別する識別手段をさらに備えることを特徴とする請求項 1 6 記載の信号記録装置。

【請求項 2 2】 上記識別手段は、候補区間数及び各候補区間の重要度が所定の閾値以上である場合に適用可能なジャンルと判定することを特徴とする請求項 2 1 記載の信号記録装置。

【請求項 2 3】 音響信号の所定時間区間毎に短時間振幅を計算する振幅計算工程と、

上記短時間振幅に基づいて上記音響信号の重要部分の候補区間を抽出する候補区間抽出工程と、

上記音響信号の所定時間区間毎に音質を定量化する音質特徴量を抽出する特徴抽出工程と、

上記短時間振幅及び上記音質特徴量に基づいて上記候補区間の重要度を計算し、該重要度に基づいて上記候補区間が上記重要部分か否かを評価する候補区間評

価工程と、

上記候補区間評価工程にて上記重要部分であると評価された候補区間の少なくとも開始位置、終了位置及び重要度を含む索引情報を生成する索引生成工程と、
上記音響信号と共に上記索引情報を記録手段に記録する記録工程と
を有することを特徴とする信号記録方法。

【請求項 24】 上記音響信号は、イベント中に発生する音響信号であることを特徴とする請求項 23 記載の信号記録方法。

【請求項 25】 所定の処理をコンピュータに実行させるプログラムにおいて

音響信号の所定時間区間毎に短時間振幅を計算する振幅計算工程と、
上記短時間振幅に基づいて上記音響信号の重要部分の候補区間を抽出する候補区間抽出工程と、

上記音響信号の所定時間区間毎に音質を定量化する音質特徴量を抽出する特徴抽出工程と、

上記短時間振幅及び上記音質特徴量に基づいて上記候補区間の重要度を計算し、該重要度に基づいて上記候補区間が上記重要部分か否かを評価する候補区間評価工程と、

上記候補区間評価工程にて上記重要部分であると評価された候補区間の少なくとも開始位置、終了位置及び重要度を含む索引情報を生成する索引生成工程と、
上記音響信号と共に上記索引情報を記録手段に記録する記録工程と
を有することを特徴とするプログラム。

【請求項 26】 上記音響信号は、イベント中に発生する音響信号であることを特徴とする請求項 25 記載のプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、イベント等における音響信号の特徴部分に対応するインデックスを生成する音響信号処理装置及びその方法、映像信号及び／又は音響信号の記録時にそのインデックスを付与し、スキップ再生や要約再生を可能とする信号記録装

置及びその方法、並びにそのような音響信号処理又は信号記録処理をコンピュータに実行させるプログラムに関する。

【0002】

【従来の技術】

放送信号や、それが録画／録音された映像／音響信号から重要なシーンを自動的に検出してインデックスを付与したり、要約映像を作成したりすることは、その内容を容易に把握したり、必要な部分を高速に検索するために有用である。例えば、スポーツイベント等の映像の場合、盛り上がった部分で自動的にインデックスを発生させ、そのインデックスを映像／音響信号に多重するなどして付与することができれば、その映像についてのダイジェスト版の作成や、2次利用の際の特定のシーンの検索が容易に可能となる。

【0003】

このような理由から、下記の特許文献1には、全周波数帯域のパワーレベルと、特定の周波数成分のパワーレベルの大小との組み合わせにより、スポーツ等のイベントの盛り上がり部分を自動的に検出し、インデックスとして付与する技術が提案されている。この技術では、観客が盛り上がった音を検出するため、多くのイベントに汎用的に適用でき、イベントの進行上重要な箇所を自動的に検出することができる。

【0004】

【特許文献1】

特開2001-143451号公報

【0005】

【発明が解決しようとする課題】

しかしながら、上述した特許文献1に記載の技術では、例えばスペクトルの形状といった音質的な要因が評価されないため、原理的に検出精度が低く、特定の周波数に他の音が混入する場合には適用することができないという問題がある。

【0006】

したがって、この技術は、イベント会場において放送局等のプロフェッショナルが録音した、他の信号がミキシングされていない音響信号にしか適用すること

ができず、アナウンスや解説、コマーシャルメッセージ等の挿入音声に加わった音響信号、例えば放送信号には適用することができない。また、観客など、アマチュアが独自に録音する場合にも、会話音声等が重畳するため、適用は困難である。

【0007】

本発明は、このような従来の実情に鑑みて提案されたものであり、イベント等における音響信号から盛り上がり部分を高精度に検出し、その盛り上がり部分に対応するインデックスを生成する音響信号処理装置及びその方法、映像信号及び／又は音響信号の記録時に盛り上がり部分に対してインデックスを付与し、スキップ再生や要約再生を可能とする信号記録装置及びその方法、並びにそのような音響信号処理又は信号記録処理をコンピュータに実行させるプログラムを提供することを目的とする。

【0008】

【課題を解決するための手段】

上述した目的を達成するために、本発明に係る音響信号処理装置及びその方法は、例えばイベント中に発生する音響信号における特徴部分を検出する際に、上記音響信号の所定時間区間毎に短時間振幅を計算し、この短時間振幅に基づいて上記特徴部分の候補区間を抽出する。また、上記音響信号の所定時間区間毎に音質を定量化する音質特徴量を抽出する。そして、上記短時間振幅及び上記音質特徴量に基づいて、上記候補区間が上記特徴部分か否かを評価する。

【0009】

ここで、本発明に係る音響信号処理装置及びその方法では、上記特徴部分の少なくとも開始位置及び終了位置を含む索引情報を生成するようにしてもよい。

【0010】

このような音響信号処理装置及びその方法は、音響信号の短時間振幅に基づいて特徴部分の候補区間を抽出し、短時間振幅及び音響特徴量に基づいてこの候補区間が特徴部分か否かを評価する。

【0011】

また、上述した目的を達成するために、本発明に係る信号記録装置及びその方

法は、例えばイベント中に発生する音響信号の所定時間区間毎に短時間振幅を計算し、上記短時間振幅に基づいて上記音響信号の重要部分の候補区間を抽出する。また、上記音響信号の所定時間区間毎に音質を定量化する音質特徴量を抽出する。そして、上記短時間振幅及び上記音質特徴量に基づいて上記候補区間の重要度を計算して該重要度に基づいて上記候補区間が上記重要部分か否かを評価し、上記重要部分であると評価された候補区間の少なくとも開始位置、終了位置及び重要度を含む索引情報を生成し、上記音響信号と共に上記索引情報を記録手段に記録する。

【0012】

このような信号記録装置及びその方法は、音響信号の短時間振幅に基づいて重要部分の候補区間を抽出し、短時間振幅及び音響特徴量に基づいてこの候補区間が重要部分か否かを評価し、重要部分である場合には、その候補区間の少なくとも開始位置、終了位置及び重要度を含む索引情報を上記音響信号と共に記録手段に記録する。

【0013】

また、本発明に係るプログラムは、上述した音響信号処理又は信号記録処理をコンピュータに実行させるものであり、本発明に係る記録媒体は、そのようなプログラムが記録されたコンピュータ読み取り可能なものである。

【0014】

【発明の実施の形態】

以下、本発明を適用した具体的な実施の形態について、図面を参照しながら詳細に説明する。

【0015】

一般に、多数のまとまった観客の集まるイベント、例えばスポーツイベント等では、注目すべき場面で多数の観客が同時に歓声、拍手や効果音を発することによる特有の音響効果（以下、「盛り上がり」という。）が生じる。

【0016】

一例として、野球放送の音響信号を図1（A）に示す。前半の平常区間ではアナウンサの解説音声が支配的であり、その背景に観客の発する音が小音量にて重

畳されている。時刻 t_2 で打者がヒットを打つと、後半の盛り上がり区間では観客の盛り上がり音が支配的になっている。ここで、平常時（時刻 t_1 ）と盛り上がり時（時刻 t_3 ）とにおける音響信号の短時間スペクトルをそれぞれ図 1（B）、（C）に示す。

【0017】

図 1（A）～（C）から分かるように、観客が盛り上がっている区間では、平常区間と比較して振幅の大きい時間が長めに持続することや、短時間スペクトルが平常区間とは異なったパターンを示すことなど、振幅や周波数構成に違いが見られる。その一方で、振幅の大きい部分は平常区間においても発生しているため、特に放送信号の場合、振幅の大小のみで盛り上がり区間か否かを判断することはできないことも確認される。

【0018】

そこで、第 1 の実施の形態における音響信号処理装置では、このような音響信号の振幅や周波数構成の違いを利用することで、スポーツ等の観客が集まるイベントにおける盛り上がり部分を重要シーンとして精度よく検出する。具体的には、振幅の大きい時間が所定時間以上持続した区間を盛り上がり区間の候補とし、各候補区間について振幅や音質に関する特徴量を用いて盛り上がりの程度を表すスコアを計算する。そして、このスコアに基づいて真の盛り上がり区間を検出する。

【0019】

この音響信号処理装置の概略構成を図 2 に示す。図 2 に示すように、音響信号処理装置 10 は、振幅計算部 11 と、挿入検出部 12 と、候補区間抽出部 13 と、特徴抽出部 14 と、候補区間評価部 15 と、索引生成部 16 とから構成されている。

【0020】

振幅計算部 11 は、入力された音響信号の所定時間区間毎に 2 乗平均や絶対値の平均を計算することで短時間振幅 $A(t)$ を計算する。なお、この振幅計算部 11 の前段に帯域通過フィルタを設けて予め不要な周波数成分を除去し、歓声等の振幅を検出するために必要十分な帯域に絞ってから短時間振幅 $A(t)$ を計算

するようにしても構わない。振幅計算部 1 1 は、計算した短時間振幅 $A(t)$ を挿入検出部 1 2 及び候補区間抽出部 1 3 に供給する。

【0 0 2 1】

挿入検出部 1 2 は、入力された音響信号が放送信号である場合に、リプレイシーンやコマーシャルメッセージ（以下、「コマーシャル」と略称する。）等の他の情報が挿入された区間を検出する。ここで、挿入シーンは長くても 1 分程度であり、挿入前後では音量が極端に小さくなるという特徴があり、一方で本編の音響信号は観客音が常に重畳されているため、音量が極端に小さくなることは殆ど生じない。そこで、挿入検出部 1 2 は、振幅計算部 1 1 から供給された短時間振幅 $A(t)$ が所定の閾値未満となる小音量区間が所定時間内に複数回発生した場合に、この小音量区間によって挟まれた区間を挿入区間として検出する。

【0 0 2 2】

また、挿入検出部 1 2 において音響信号の他に映像信号も利用できる場合には、民間放送のコマーシャルをより正確に検出するために、例えば本件発明者が先に提案している特開 2 0 0 2 - 1 6 8 7 3 号公報に記載された技術を用いることができる。この技術は、以下のように要約される。

【0 0 2 3】

すなわち、日本国内においては、特殊な例を除くほぼ全てのコマーシャルが 15 秒、30 秒、60 秒の長さで製作されており、また、各コマーシャルの前後では必然的に音量が下がり、映像信号が切り替わるため、これらを検出の「必須条件」とする。また、規格に基づいて製作されるという制約、短い時間で宣伝効果を上げるためのものであるという制約、及び番組構成上の都合などによる制約の元で製作された結果として現れる一定の傾向を示す特徴を検出の「付加条件」とする。さらに、上記付加条件を満たす区間がオーバーラップして存在する場合には、少なくともその何れかの区間は正しいコマーシャル区間では有り得ないという条件を検出の「論理条件」とする。そして、「必須条件」に基づき決定論的にコマーシャル候補を抽出し、「付加条件」に基づくコマーシャルらしさの統計論的な評価により候補を選択し、「論理条件」により候補のオーバーラップ関係を解消することにより、高精度にコマーシャルを検出することができる。

【0024】

挿入検出部12は、以上のようにして検出した挿入区間に関する挿入区間情報を候補区間評価部15に供給する。なお、入力された音響信号が放送信号でない場合には、この挿入検出部12を設けなくても構わない。

【0025】

候補区間抽出部13は、振幅計算部11から供給された短時間振幅 $A(t)$ を用いて、盛り上がり区間の候補を抽出する。上述したように、盛り上がり区間では、平均的に音量の大きい区間が長く継続するため、候補区間抽出部13は、予め振幅閾値 A_{thsd} と時間閾値 T_{thsd} とを設定しておき、 $A(t) \geq A_{thsd}$ である区間長 T が $T \geq T_{thsd}$ である場合に、その区間を盛り上がり区間の候補とし、その開始位置及び終了位置を抽出する。なお、振幅閾値 A_{thsd} としては、予め所定値を設定してもよく、対象とする音響信号の振幅の平均及び分散に基づいて設定してもよい。前者の場合、放送に合わせて実時間で処理することができ、後者の場合、スタジアム、放送局、ミキシング、或いはイベント毎の音量の差に関して正規化された閾値を設定することができる。

【0026】

ここで、図3(A)に示す図1(A)と同一の音響信号について、2つの候補区間が抽出された例を図3(B)に示す。図3(B)に示すように、第1の候補区間は平常区間に属し、第1の候補区間は盛り上がり区間に属している。

【0027】

特徴抽出部14は、入力された音響信号から音質に関する音質特徴量 X を抽出し、盛り上がり時に特有の音質を定量化する。具体的には、図4のフローチャートに示すように、始めにステップS1において所定時間区間の音響信号を取得し、ステップS2において短時間フーリエ変換やLPC(Linear Predictive Coding)法などを用いてパワースペクトラム係数 S_0, \dots, S_{M-1} に変換する。なお、 M はスペクトラムの次数を示す。

【0028】

続いてステップS3において、以下の式(1)に従ってスペクトラム係数 S_0, \dots, S_{M-1} の荷重和を計算し、音質特徴量 X を得る。

【0029】

【数1】

$$X = \sum_{m=0}^{M-1} W_m S_m + \theta \quad \cdots (1)$$

【0030】

ここで、式（1）において W_m は荷重係数を示し、 θ は所定のバイアス値を示す。この荷重係数 W_m 及びバイアス値 θ の決定には種々の統計的判別法を利用することができる。例えば、予め多数のシーンの盛り上がり程度を主観的に評価し、スペクトラム係数と望ましい特徴量値（例えば0.0～1.0）とを組にした学習サンプルを用意して重回帰分析により直線近似される荷重を求め、荷重係数 W_m 及びバイアス値 θ を決定することができる。また、パーセプトロン等のニューラルネットの技法を用いてもよく、ベイズ判別法やベクトル量子化等の判別法を用いてもよい。

【0031】

なお、特徴抽出部14では、パワースペクトラム係数に代えて、対数ケプストラム係数の逆フーリエ変換であるケプストラム係数や、スペクトラム係数の固有ベクトルによる変換であるカルーネンレーベ（Karhunen-Loeve；KL）変換を用いてもよい。これらはスペクトラムの概形を集約する変換であるため、それぞれ低次の項のみを用いて荷重和を計算して音質特徴量 X とすればよい。

【0032】

図3（A）に示す音響信号について、KL変換を用いて計算した音質特徴量 X を図3（C）に示す。図3（C）に示すように、平常区間に属する第1の候補区間と盛り上がり区間に属する第2の候補区間とでは、音質特徴量 X の値に明確な違いが現れている。

【0033】

図2に戻って、候補区間評価部15は、振幅計算部11から供給された短時間振幅 $A(t)$ と特徴抽出部14から供給された音質特徴量 X とに基づいて、各候

補区間における盛り上がりの程度をスコアとして定量化する。具体的には、図 5 のフローチャートに示すように、始めにステップ S 1 0 において候補区間抽出部 1 3 から候補区間を取得し、ステップ S 1 1 において、短時間振幅 $A(t)$ と音質特徴量 X とを用いて、各候補区間の区間長 y_1 、短時間振幅 $A(t)$ の最大値 y_2 、短時間振幅 $A(t)$ の平均値 y_3 、音質特徴量 X が所定の閾値を超える長さ y_4 、音質特徴量 X の最大値 y_5 、音質特徴量 X の平均値 y_6 を計算し、候補区間の特徴ベクトルとする。なお、音質特徴量 X が所定の閾値を超える長さ y_4 の代わりに、音質特徴量 X が所定の閾値を超える長さの候補区間長に対する割合を特徴量として用いても構わない。

【0 0 3 4】

ここで、候補区間評価部 1 5 において候補区間内での音質特徴量 X の最大値を計算することで、盛り上がり時にアナウンサの音声为重畳され、スペクトル分布が歪んでいるような場合でも、瞬時的にアナウンサの音声途切れる部分の特徴量を用いることができ、放送信号のような他の音声为重畳されている音響信号についても適用可能となる。

【0 0 3 5】

続いてステップ S 1 2 において、以下の式 (2) に従って各候補区間の盛り上がりの程度を表すスコア値 z を計算する。

【0 0 3 6】

【数 2】

$$z = \sum_{i=1}^6 u_i y_i + u_0 \quad \cdots (2)$$

【0 0 3 7】

ここで、式 (2) において u_i は荷重係数を示し、 u_0 は所定のバイアス値を示す。この荷重係数 u_i 及びバイアス値 u_0 の決定には種々の統計的判別法を利用することができる。例えば、予め多数のシーンの盛り上がり程度を主観的に評価し、特徴ベクトルと望ましいスコア値 (例えば 0.0 ～ 1.0) とを組にした

学習サンプルを用意して重回帰分析により直線近似される荷重を求め、荷重係数 u_i 及びバイアス値 u_0 を決定することができる。また、パーセプトロン等のニューラルネットの技法を用いてもよく、ベイズ判別法やベクトル量子化等の判別法を用いてもよい。

【0038】

最後にステップ S13 において、候補区間のうち、挿入検出部 12 から供給された挿入区間情報で表される区間、或いはスコア値 z が所定の閾値以下である候補区間を盛り上がり区間から除外する。

【0039】

候補区間評価部 15 は、以上のようにして求めた盛り上がり区間の開始位置及び終了位置とスコア値 z とを索引生成部 16 に供給する。

【0040】

索引生成部 16 は、候補区間評価部 15 から供給された盛り上がり区間の開始位置及び終了位置とスコア値 z とを含むインデックスを生成し、そのインデックスを出力する。この際、予め抽出区間数を設定し、その抽出区間数に達するまでスコア値 z の大きい順にインデックスを生成するようにしても構わない。また、予め抽出時間長を設定し、その抽出時間長に達するまでスコア値 z の大きい順にインデックスを生成するようにしても構わない。

【0041】

以上のように、第 1 の実施の形態における音響信号処理装置 10 によれば、短時間振幅 $A(t)$ が振幅閾値 A_{thsd} 以上である区間長 T が時間閾値 T_{thsd} 以上である区間を候補区間とし、各候補区間について振幅や音質に関する特徴量 ($y_1 \sim y_6$) を用いて盛り上がりの程度を表すスコア値 z を計算することで、真の盛り上がり区間を高精度に検出し、その盛り上がり区間に対応するインデックスを生成することができる。

【0042】

次に、第 2 の実施の形態における記録再生装置は、上述した音響信号処理装置 10 が適用されたものであり、放送信号の録画／録音時に盛り上がり区間の開始位置及び終了位置等を記録し、再生時に盛り上がり区間のみをスキップ再生した

り、要約映像を再生したりすることを可能とする。

【0043】

この記録再生装置の概略構成を図6に示す。図6に示すように、記録再生装置20は、受信部21と、ジャンル選択部22と、記録部23と、重要シーン検出部24と、サムネイル生成部25と、シーン選択部26と、再生部27とから構成されている。ここで、重要シーン検出部24は、上述した音響信号処理装置10に対応する。

【0044】

この記録再生装置20における映像／音響信号の記録動作について、図6と図7のフローチャートとを用いて説明する。始めにステップS20において、受信部21は、図示しないタイマやユーザからの指示により放送信号を受信して復調し、復調した放送信号を記録部23に対して記録する録画／録音動作を開始する。

【0045】

次にステップS21において、ジャンル選択部22は、電子番組ガイド (Electronic Program Guide; EPG) の情報を用いて、放送信号のジャンルが観客の歓声等を伴うイベントに関するものであるか否かを判別する。ステップS21において観客の歓声等を伴うイベントに関するものでない場合 (No) にはステップS22に進み、受信部21は、記録部23に対する通常の録画／録音動作を行い、ステップS23でタイマ又はユーザの指示により録画／録音を終了する。一方、観客の歓声等を伴うイベントに関するものである場合 (Yes) には、ステップS24に進む。

【0046】

なお、ステップS21では、EPGの情報を用いるほか、ユーザがジャンルを指示するようにしてもよい。また、放送信号からジャンルを自動推定しても構わない。例えば、候補区間数及び各候補区間のスコア値 z が所定の閾値以上である場合に、適用可能なジャンルと判定することができる。

【0047】

ステップS24では、受信部21は、記録部23に対する通常の録画／録音動

作を行うと同時に、重要シーン検出部 24 は、重要シーンとして盛り上がり区間の開始位置及び終了位置を検出する。

【0048】

続いてステップ S 25 において、受信部 21 は、タイマ又はユーザの指示により録画／録音を終了するが、引き続きステップ S 26 において、重要シーン検出部 24 は、盛り上がり区間の開始位置及び終了位置、スコア値 z を含むインデックスを記録部 23 に記録し、サムネイル生成部 25 は、盛り上がり区間のサムネイル画像を記録部 23 に記録する。

【0049】

このように、記録再生装置 20 では、放送信号の録画／録音時に盛り上がり区間を検出し、その盛り上がり区間の開始位置及び終了位置、スコア値 z を含むインデックスや、盛り上がり区間のサムネイル画像を記録部 23 に記録する。したがって、記録再生装置 20 は、記録部 23 に記録されたサムネイル画像を表示することができるほか、盛り上がり区間のインデックスを利用して、以下に説明するようなスキップ再生や要約再生が可能となる。

【0050】

この記録再生装置 20 におけるスキップ再生動作について、図 6 と図 8 のフローチャートとを用いて説明する。始めにステップ S 30 において、再生部 27 は、ユーザからの指示により記録部 23 に記録されている映像／音響信号の再生を開始し、ステップ S 31 において、ユーザから停止命令がなされたか否かを判別する。ステップ S 31 において停止命令がなされた場合 (Yes) には再生動作を終了し、そうでない場合 (No) にはステップ S 32 に進む。

【0051】

ステップ S 32 では、再生部 27 は、ユーザからスキップ命令がなされたか否かを判別する。スキップ命令がなされていない場合 (No) にはステップ S 30 に戻って再生動作を継続し、スキップ命令がなされた場合 (Yes) にはステップ S 33 に進む。

【0052】

ステップ S 33 では、再生部 27 は、盛り上がり区間に付与されたインデック

スを参照して、次の開始インデックス点への頭出しを行ってステップ S 3 0 に戻る。

【 0 0 5 3 】

続いて、この記録再生装置 2 0 における要約再生動作について、図 6 と図 9 のフローチャートとを用いて説明する。始めにステップ S 4 0 において、シーン選択部 2 6 は、例えば所定の時間長に合わせて、再生するシーンをスコア値 z に基づいて選択し、その開始位置及び終了位置を決定する。

【 0 0 5 4 】

次にステップ S 4 1 において、再生部 2 7 は、最初の開始インデックス点への頭出しを行い、ステップ S 4 2 において再生動作を行う。

【 0 0 5 5 】

続いてステップ S 4 3 において、再生部 2 7 は、終了インデックス点まで再生したか否かを判別する。終了インデックス点まで再生していない場合 (No) には、ステップ S 4 2 に戻って再生動作を継続し、終了インデックス点まで再生した場合 (Yes) にはステップ S 4 4 に進む。

【 0 0 5 6 】

ステップ S 4 4 では、再生部 2 7 は、次の開始インデックス点があるか否かを判別する。次の開始インデックス点がある場合 (Yes) には、ステップ S 4 5 において、再生部 2 7 は、その開始インデックス点への頭出しを行ってステップ S 4 2 に戻り、次の開始インデックス点がない場合 (No) には、再生動作を終了する。

【 0 0 5 7 】

以上のように、本実施の形態における記録再生装置 2 0 によれば、放送信号の録画／録音時に盛り上がり区間を検出し、その盛り上がり区間の開始位置及び終了位置、スコア値 z を含むインデックスや、盛り上がり区間のサムネイル画像を記録部 2 3 に記録することにより、サムネイル画像を例えば重要度を示すスコア値 z に応じて表示することができるほか、盛り上がり区間のインデックスを利用してスキップ再生や要約再生が可能となる。

【 0 0 5 8 】

なお、本発明は上述した実施の形態のみに限定されるものではなく、本発明の要旨を逸脱しない範囲において種々の変更が可能であることは勿論である。

【0059】

例えば、上述した第2の実施の形態では、映像信号と音響信号との両方を用いるものとして説明したが、これに限定されるものではなく、音響信号のみであっても同様の効果が得られる。

【0060】

また、上述の実施の形態では、ハードウェアの構成として説明したが、これに限定されるものではなく、任意の処理を、CPU (Central Processing Unit) にコンピュータプログラムを実行させることにより実現することも可能である。この場合、コンピュータプログラムは、記録媒体に記録して提供することも可能であり、また、インターネットその他の伝送媒体を介して伝送することにより提供することも可能である。

【0061】

【発明の効果】

以上詳細に説明したように本発明に係る音響信号処理装置及びその方法は、例えばイベント中に発生する音響信号における特徴部分を検出する際に、上記音響信号の所定時間区間毎に短時間振幅を計算し、この短時間振幅に基づいて上記特徴部分の候補区間を抽出する。また、上記音響信号の所定時間区間毎に音質を定量化する音質特徴量を抽出する。そして、上記短時間振幅及び上記音質特徴量に基づいて、上記候補区間が上記特徴部分か否かを評価する。

【0062】

ここで、本発明に係る音響信号処理装置及びその方法では、上記特徴部分の少なくとも開始位置及び終了位置を含む索引情報を生成するようにしてもよい。

【0063】

このような音響信号処理装置及びその方法によれば、音響信号の短時間振幅に基づいて特徴部分の候補区間を抽出し、短時間振幅及び音響特徴量に基づいてこの候補区間が特徴部分か否かを評価することにより、例えばイベントにおける盛り上がり部分を高精度に検出することができる。

【0064】

また、本発明に係る信号記録装置及びその方法は、例えばイベント中に発生する音響信号の所定時間区間毎に短時間振幅を計算し、上記短時間振幅に基づいて上記音響信号の重要部分の候補区間を抽出する。また、上記音響信号の所定時間区間毎に音質を定量化する音質特徴量を抽出する。そして、上記短時間振幅及び上記音質特徴量に基づいて上記候補区間の重要度を計算して該重要度に基づいて上記候補区間が上記重要部分か否かを評価し、上記重要部分であると評価された候補区間の少なくとも開始位置、終了位置及び重要度を含む索引情報を生成し、上記音響信号と共に上記索引情報を記録手段に記録する。

【0065】

このような信号記録装置及びその方法によれば、音響信号の短時間振幅に基づいて重要部分の候補区間を抽出し、短時間振幅及び音響特徴量に基づいてこの候補区間が重要部分か否かを評価し、重要部分である場合には、その候補区間の少なくとも開始位置、終了位置及び重要度を含む索引情報を上記音響信号と共に記録手段に記録することにより、例えば重要部分のみをスキップ再生することや、重要部分のみの要約映像を再生することが可能となる。

【0066】

また、本発明に係るプログラムは、上述した音響信号処理又は信号記録処理をコンピュータに実行させるものであり、本発明に係る記録媒体は、そのようなプログラムが記録されたコンピュータ読み取り可能なものである。

【0067】

このようなプログラム及び記録媒体によれば、上述した音響信号処理及び信号記録処理をソフトウェアにより実現することができる。

【図面の簡単な説明】**【図1】**

イベントにおける音響信号の例を示す図であり、同図（A）は、野球放送の音響信号を示し、同図（B）、（C）は、それぞれ平常時と盛り上がり時とにおける音響信号の短時間スペクトルを示す。

【図2】

第 1 の実施の形態における音響信号処理装置の概略構成を説明する図である。

【図 3】

同音響信号処理装置の候補区間抽出部及び特徴抽出部における処理の一例を説明する図であり、同図（A）は、イベントにおける音響信号の例を示し、同図（B）は、候補区間抽出部において検出される候補区間を示し、同図（C）は、特徴抽出部において計算される音質特徴量を示す。

【図 4】

同音響信号処理装置の特徴抽出部における動作を説明するフローチャートである。

【図 5】

同音響信号処理装置の候補区間評価部における動作を説明するフローチャートである。

【図 6】

第 2 の実施の形態における記録再生装置の概略構成を説明する図である。

【図 7】

同記録再生装置における録画／録音動作を説明するフローチャートである。

【図 8】

同記録再生装置におけるスキップ再生動作を説明するフローチャートである。

【図 9】

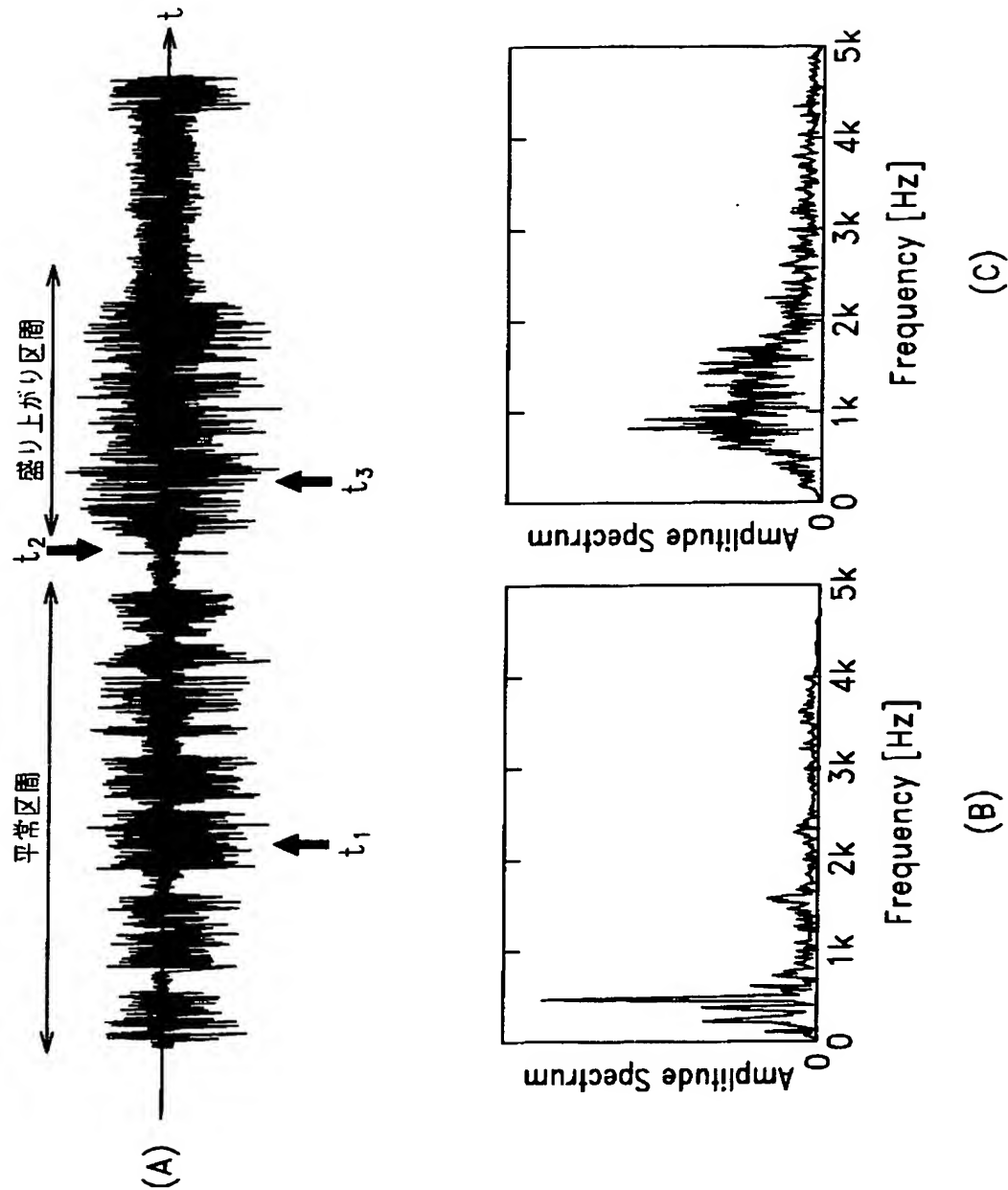
同記録再生装置における要約再生動作を説明するフローチャートである。

【符号の説明】

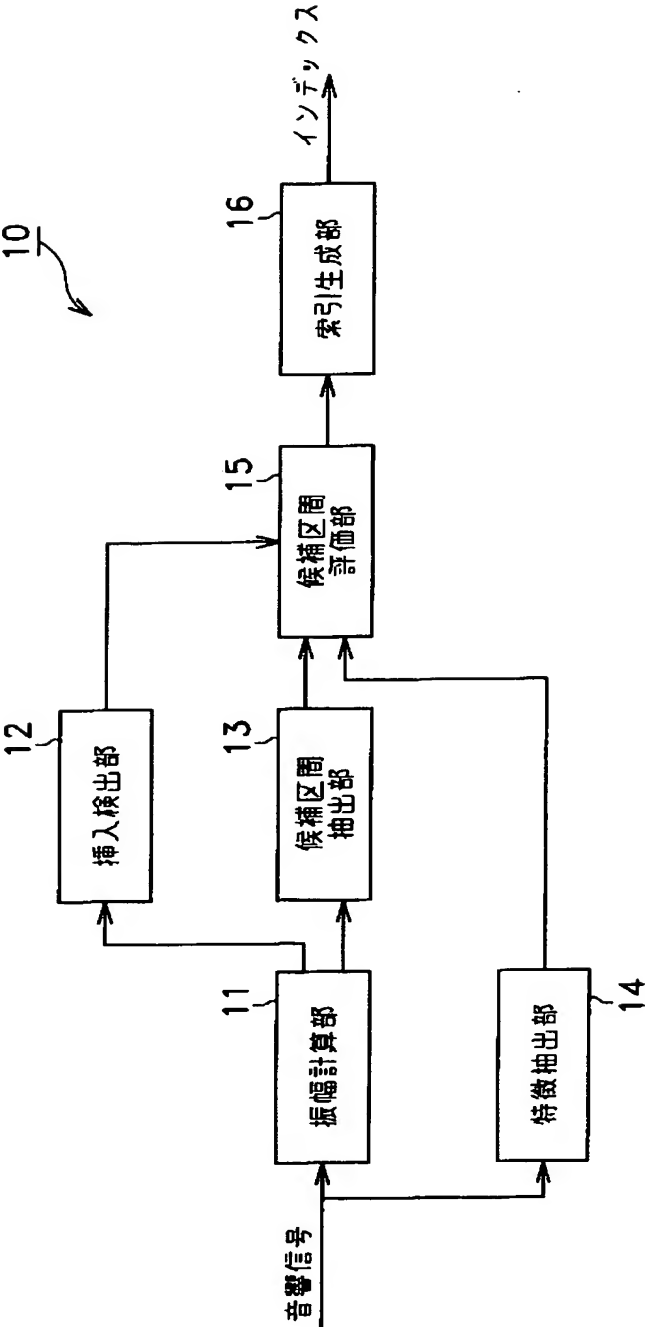
1 0 音響信号処理装置、1 1 振幅計算部、1 2 挿入検出部、1 3 候補区間抽出部、1 4 特徴抽出部、1 5 候補区間評価部、1 6 索引生成部、2 0 記録再生装置、2 1 受信部、2 2 ジャンル選択部、2 3 記録部、2 4 重要シーン検出部、2 5 サムネイル生成部、2 6 シーン選択部、2 7 再生部

【書類名】 図面

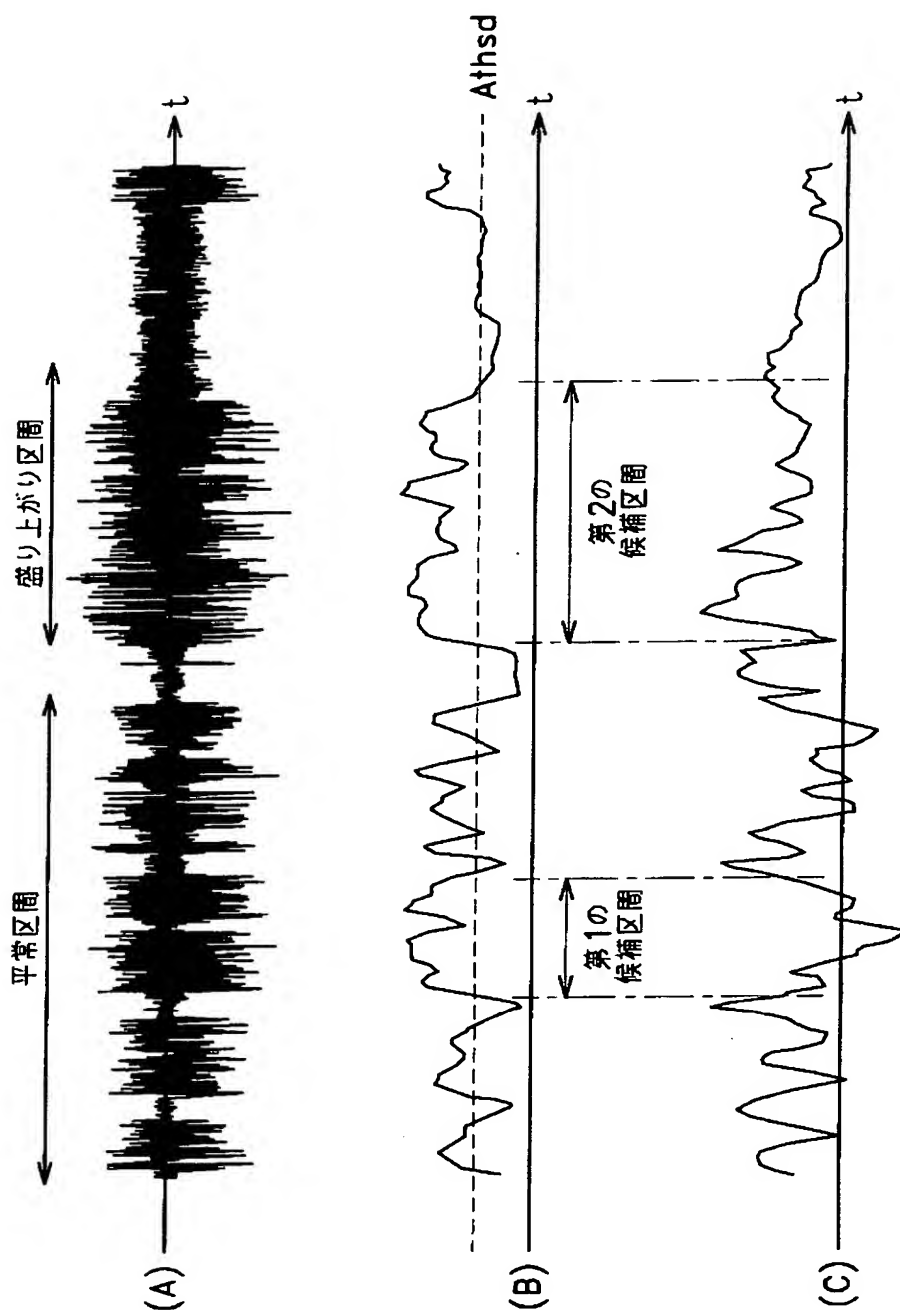
【図 1】



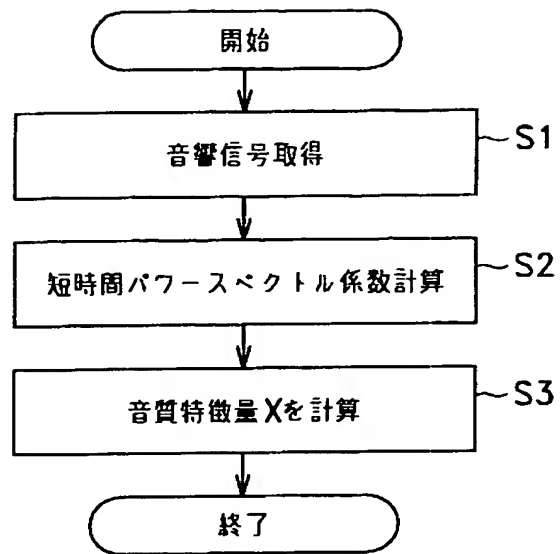
【図 2】



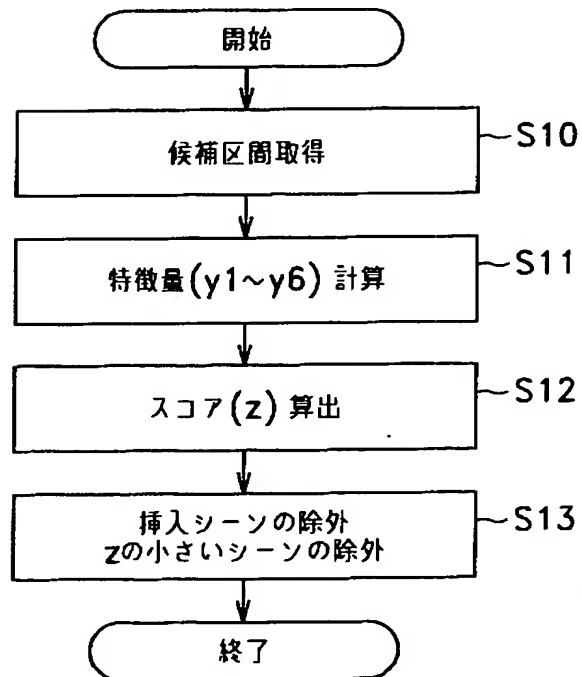
【図 3】



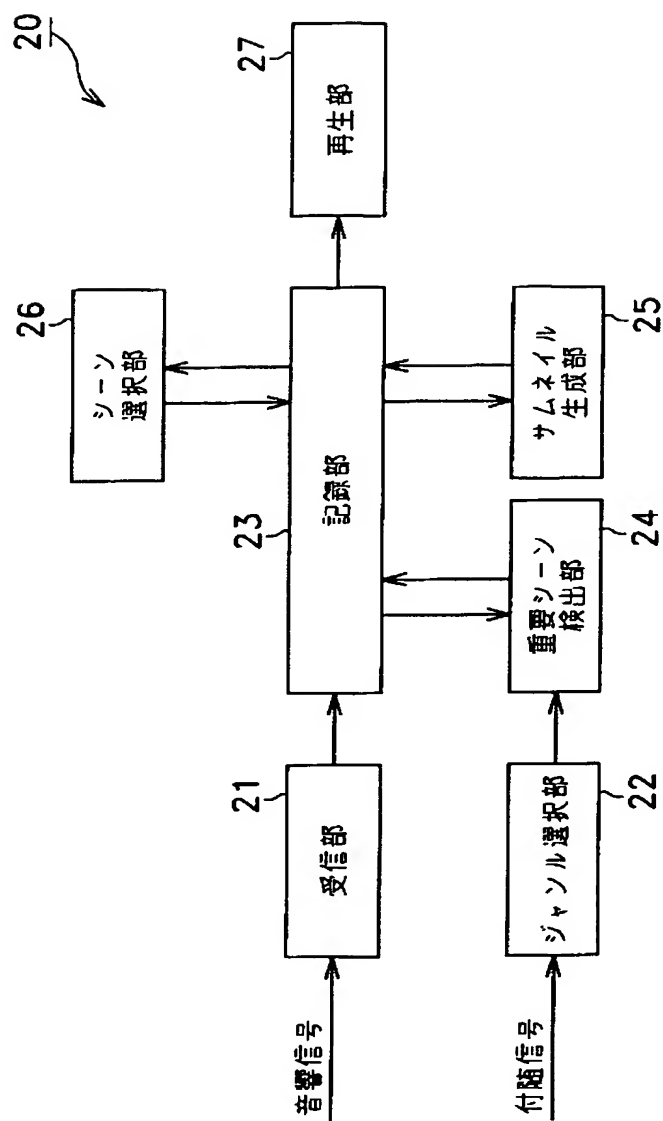
【図 4】



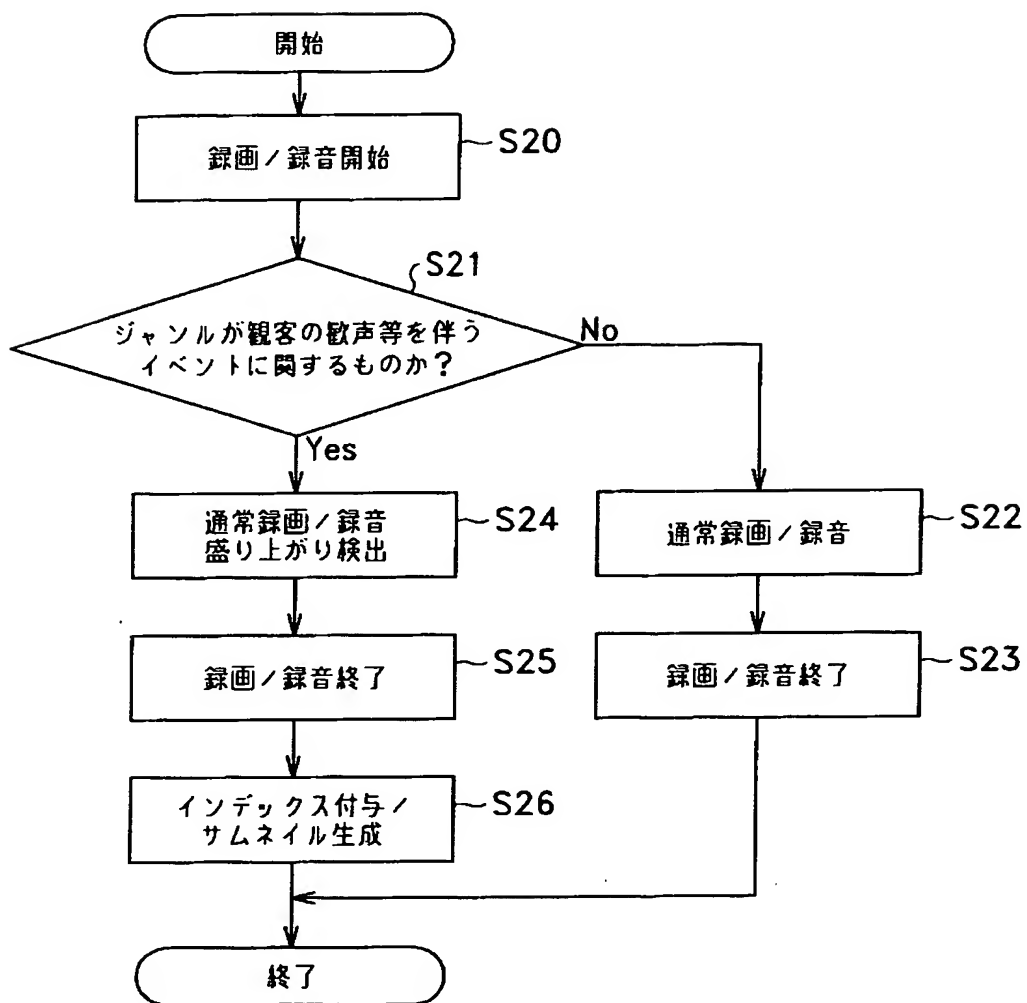
【図 5】



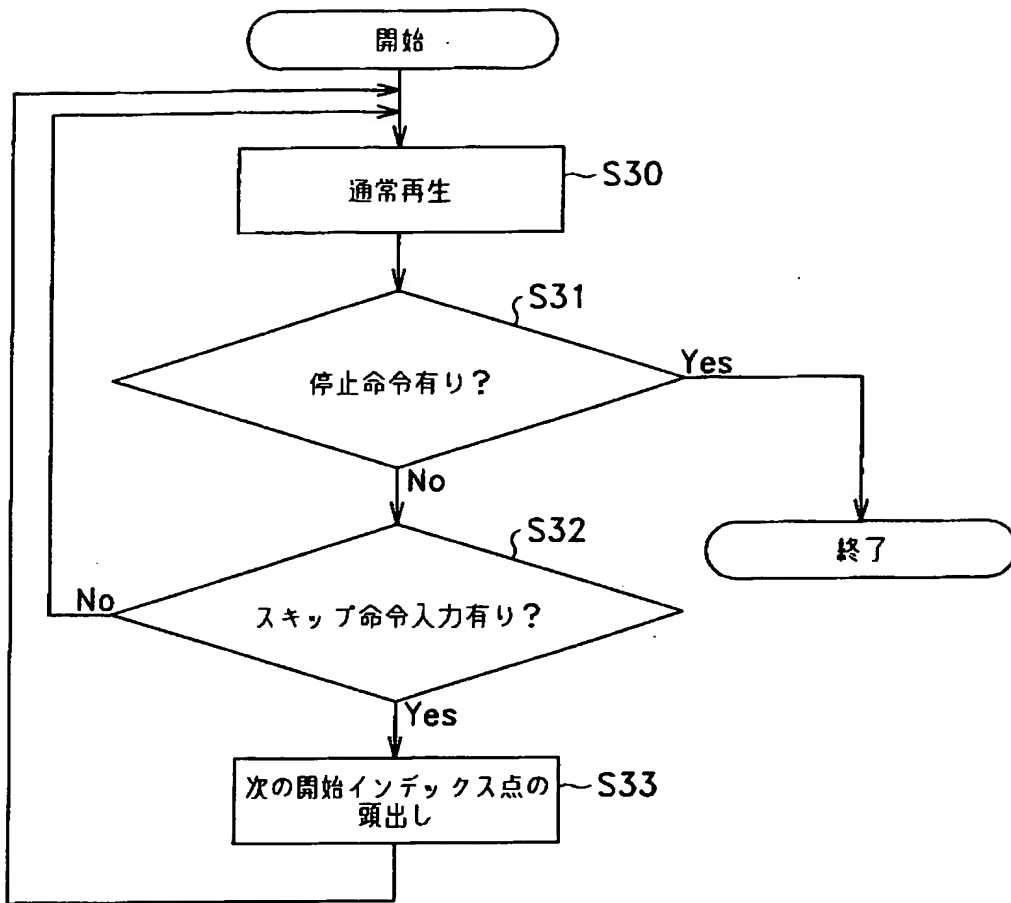
【図 6】



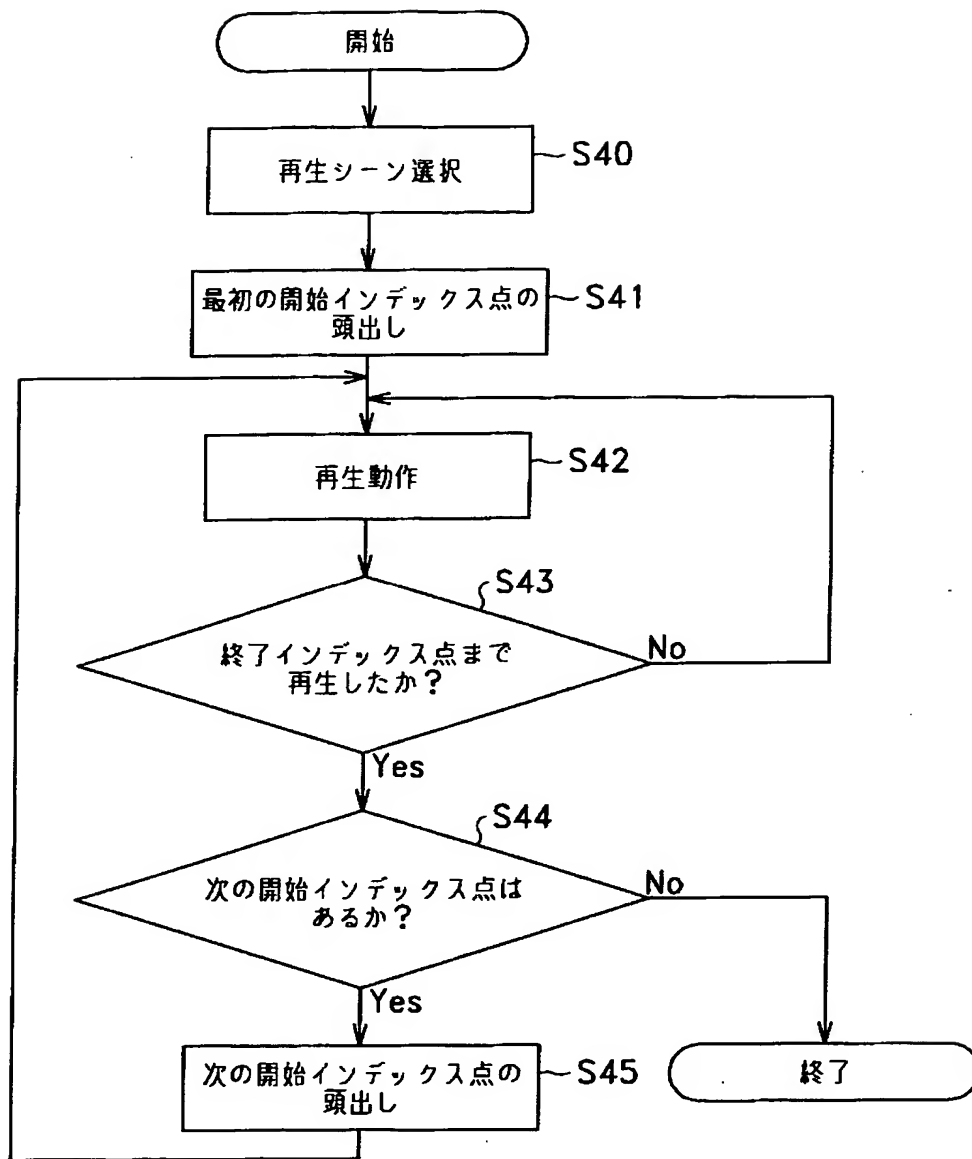
【図 7】



【図 8】



【図 9】



【書類名】 要約書

【要約】

【課題】 イベント等における音響信号から盛り上がり部分を高精度に検出し、その盛り上がり部分にインデックスを付加する。

【解決手段】 音響信号処理装置 1 0 において、候補区間抽出部 1 3 は、振幅計算部 1 1 で計算された短時間振幅が振幅閾値以上である区間長が時間閾値以上である区間を候補区間とする。また、特徴抽出部 1 4 は、音響信号から音質に関する音質特徴量を抽出し、盛り上がり時に特有の音質を定量化する。候補区間評価部 1 5 は、各候補区間について振幅や音質に関する特徴量を用いて盛り上がりの程度を表すスコア値を計算し、このスコア値に基づいて真の盛り上がり区間を検出する。そして、索引生成部 1 6 は、盛り上がり区間の開始位置及び終了位置、スコア値を含むインデックスを生成して出力する。

【選択図】 図 2

特願 2 0 0 2 - 3 6 1 3 0 2

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 2 1 8 5]

1. 変更年月日

1 9 9 0 年 8 月 3 0 日

[変更理由]

新規登録

住 所

東京都品川区北品川 6 丁目 7 番 3 5 号

氏 名

ソニー株式会社